1 Locally Weighted Linear Regression

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)}(y^{(i)} - \theta^T x^{(i)})^2$$

In the matrix notation error function can be written as :

$$J(\theta) = \frac{1}{2}(X\theta - Y)^T W(X\theta - Y)$$

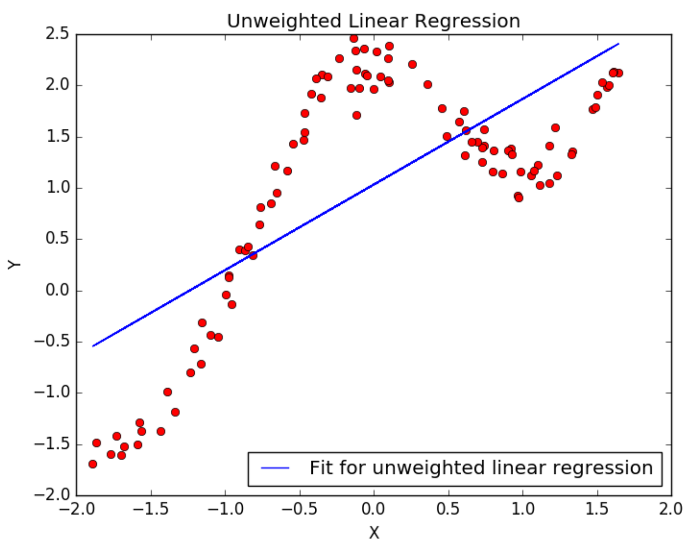
where W is diagonal matrix such that :

$$W(i,i) = exp(-\frac{(x^i - x)^2}{2\tau^2})$$

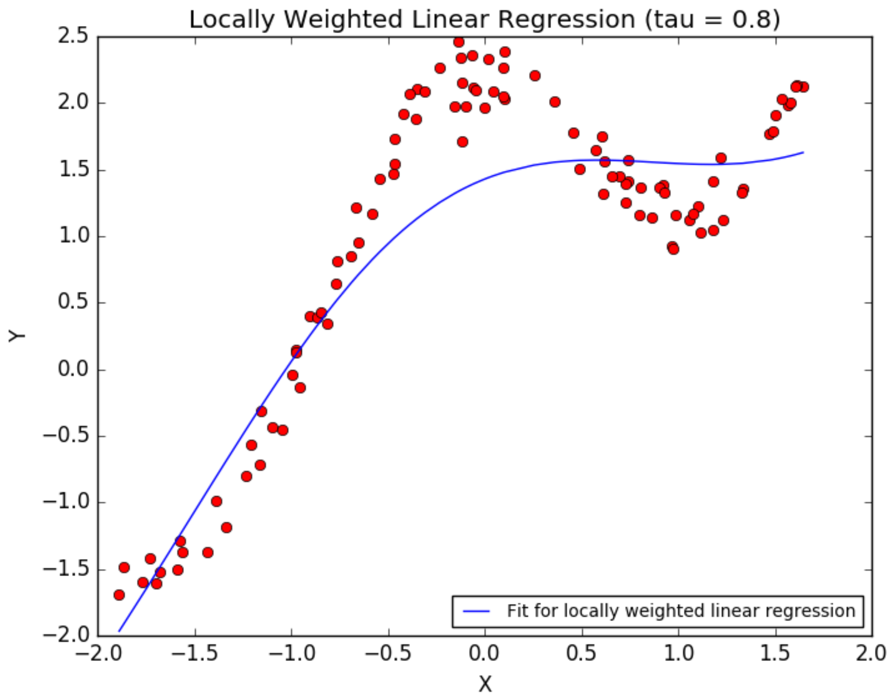
Value of θ that minimizes J(θ):

$$\theta = (X^T W X)^{-1} X^T W Y$$

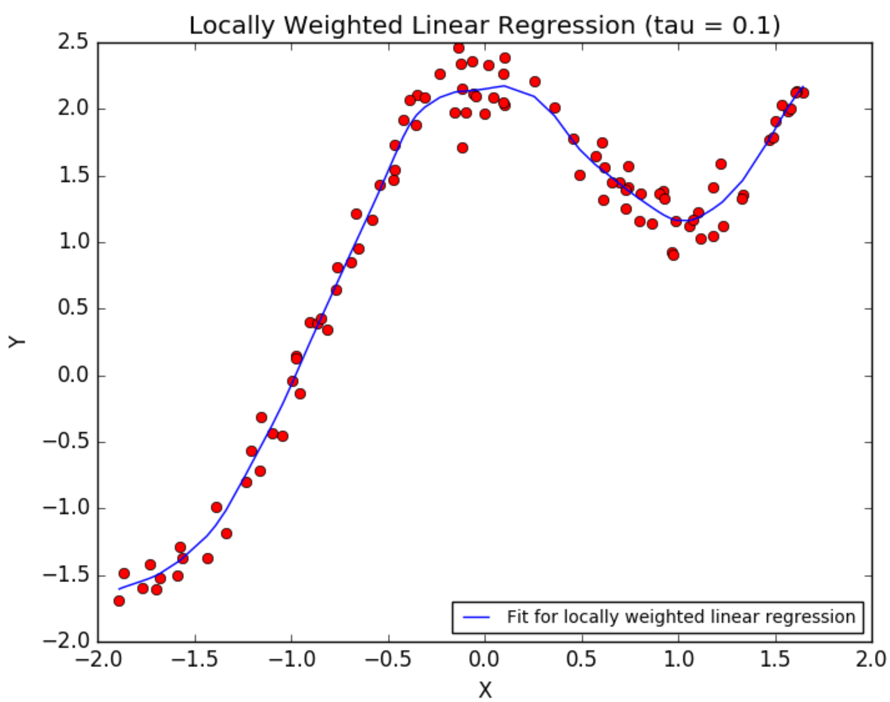
a)

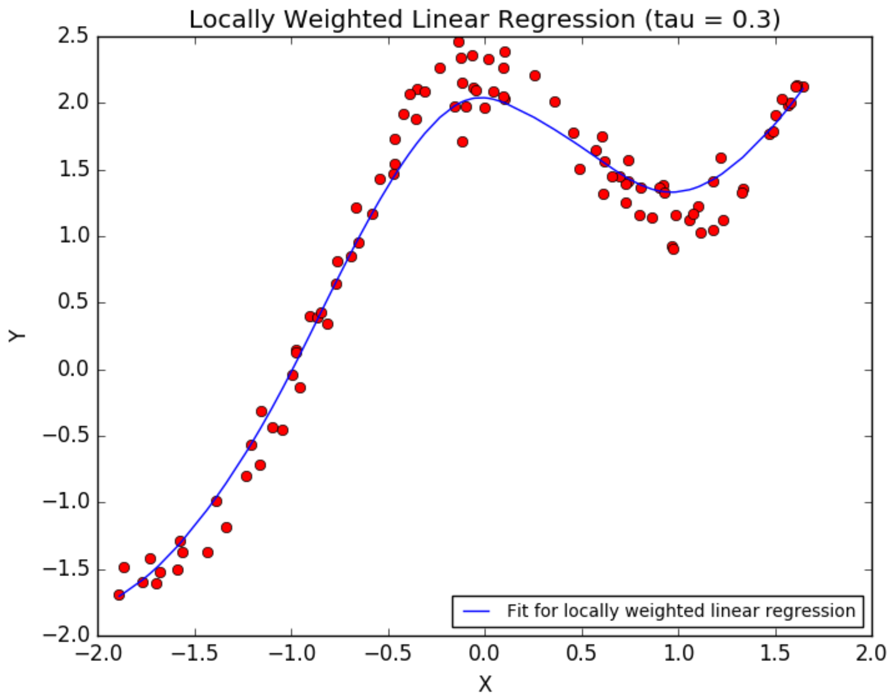


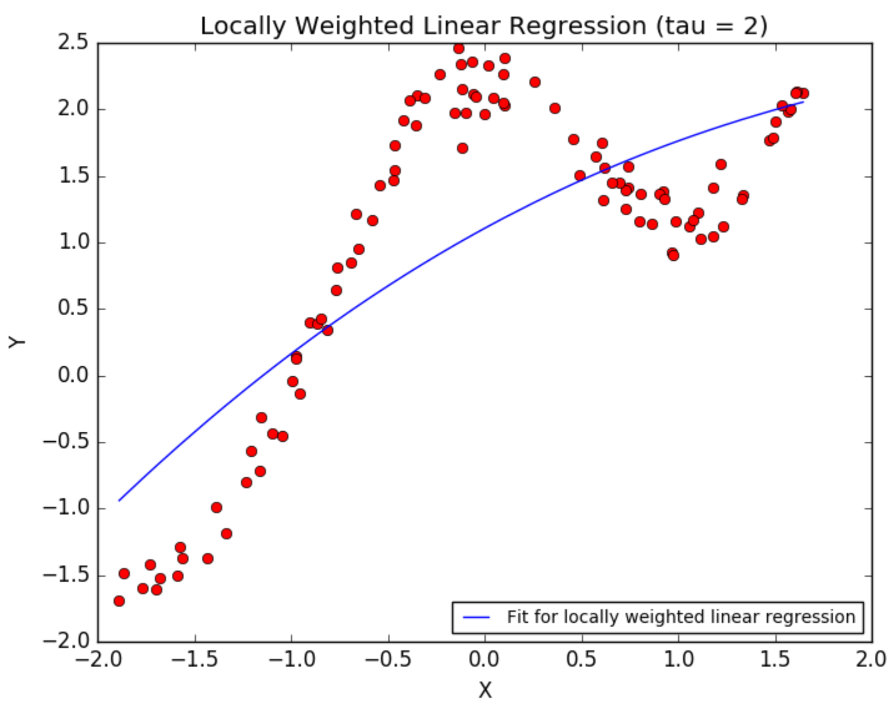
b)

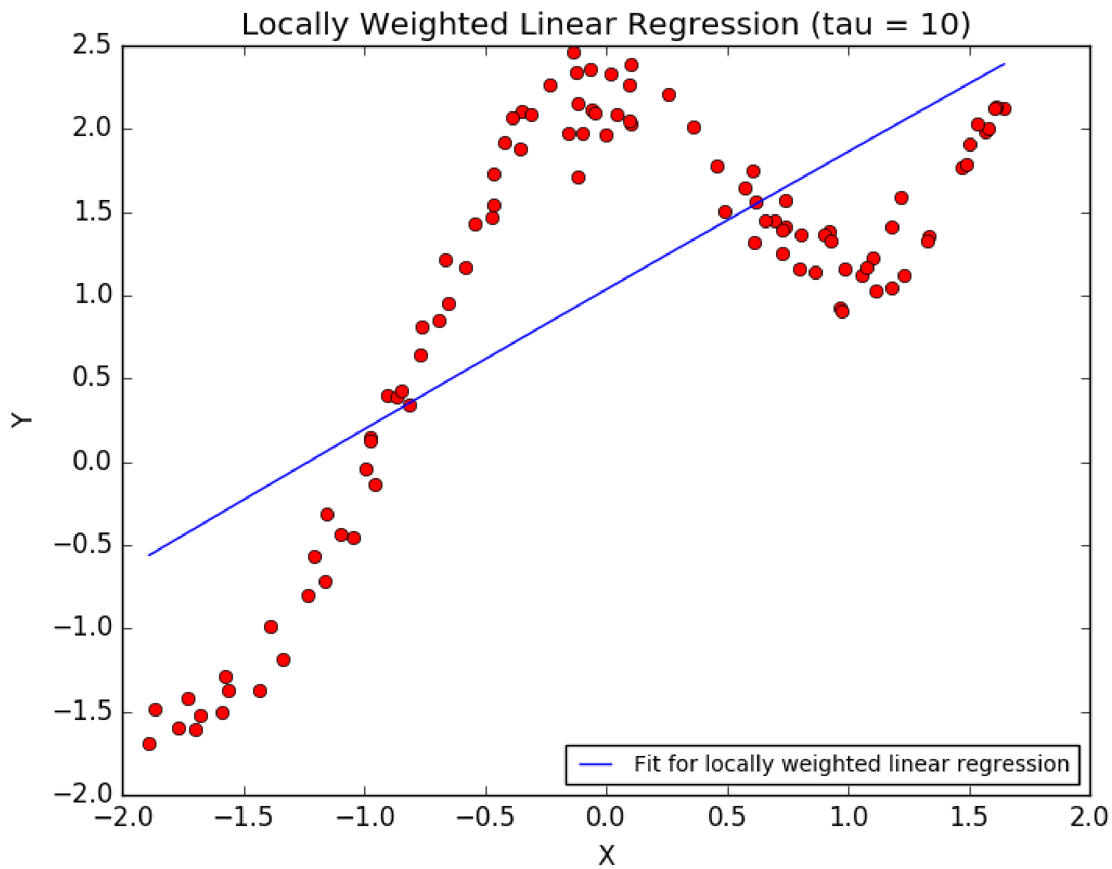


c) Comparative study of different values of τ




Locally Weighted Linear Regression (tau = 0.3)
Y
X
Fit for locally weighted linear regression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Locally Weighted Linear Regression (tau = 2)
Y
X
Fit for locally weighted linear regression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0



I think $\tau = 0.3$ works best as it is able to generalize the pattern in the data quite well without overfitting.
The parameter τ controls how quickly the weight of a training example falls off with distance of its x^i from the query point x If τ is too large weight of all training examples would be same leading to the situation same as Linear Regression and hence we would obtain a straight line.
if τ is too low the weight of most of the training examples would be close to zero and hence the curve resulting from the fit will pass through each point resulting in overfitting.

The End.